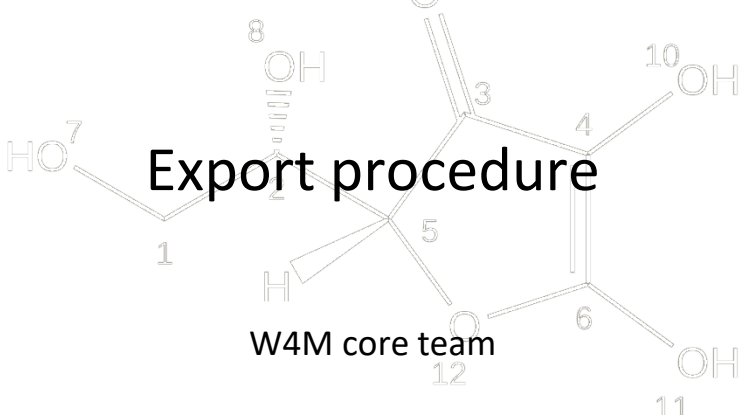
W4M



Export procedure

W4M core team

01/07/2020

v 1.0.3

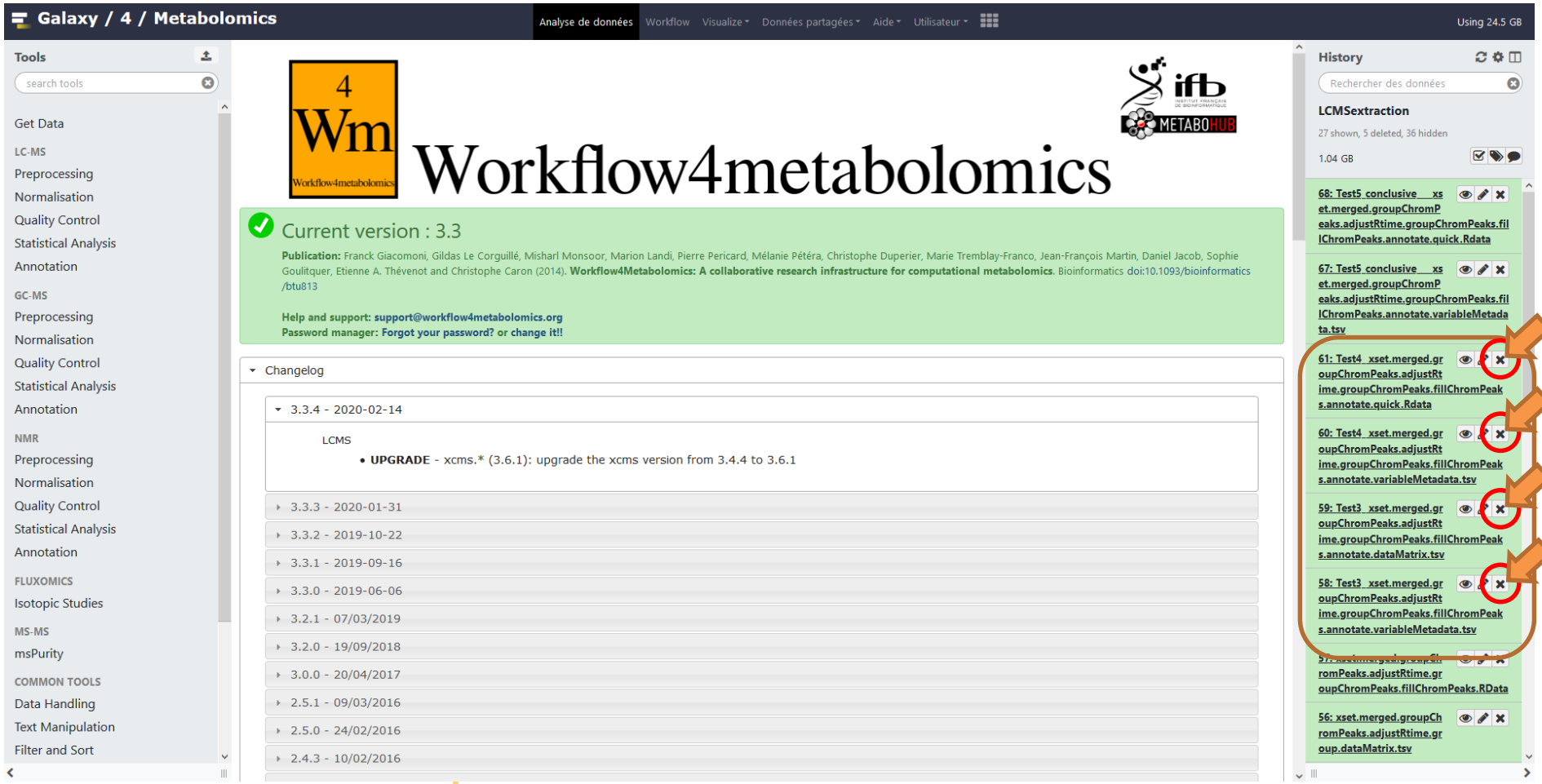
How to download your data from the ‘old’ Galaxy

https://galaxy.workflow4metabolomics.org/

GETTING YOUR DATA BACK

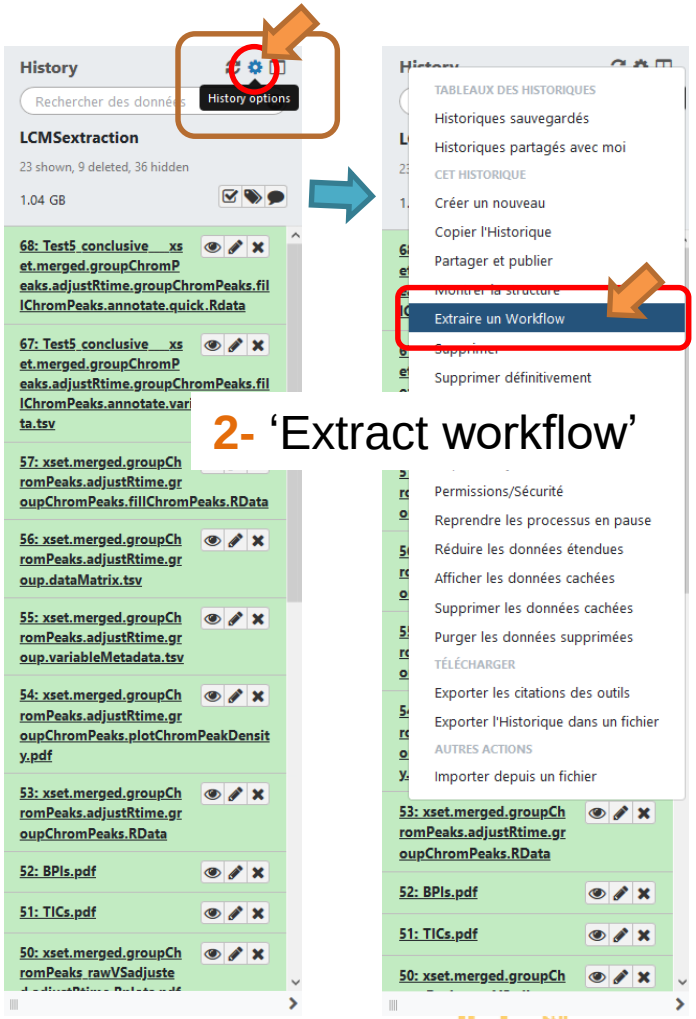
Step 1: delete unnecessary datasets

Delete datasets not needed in your workflow (*e.g.* wrong parameters)



Step 2: Extract the workflow from your history

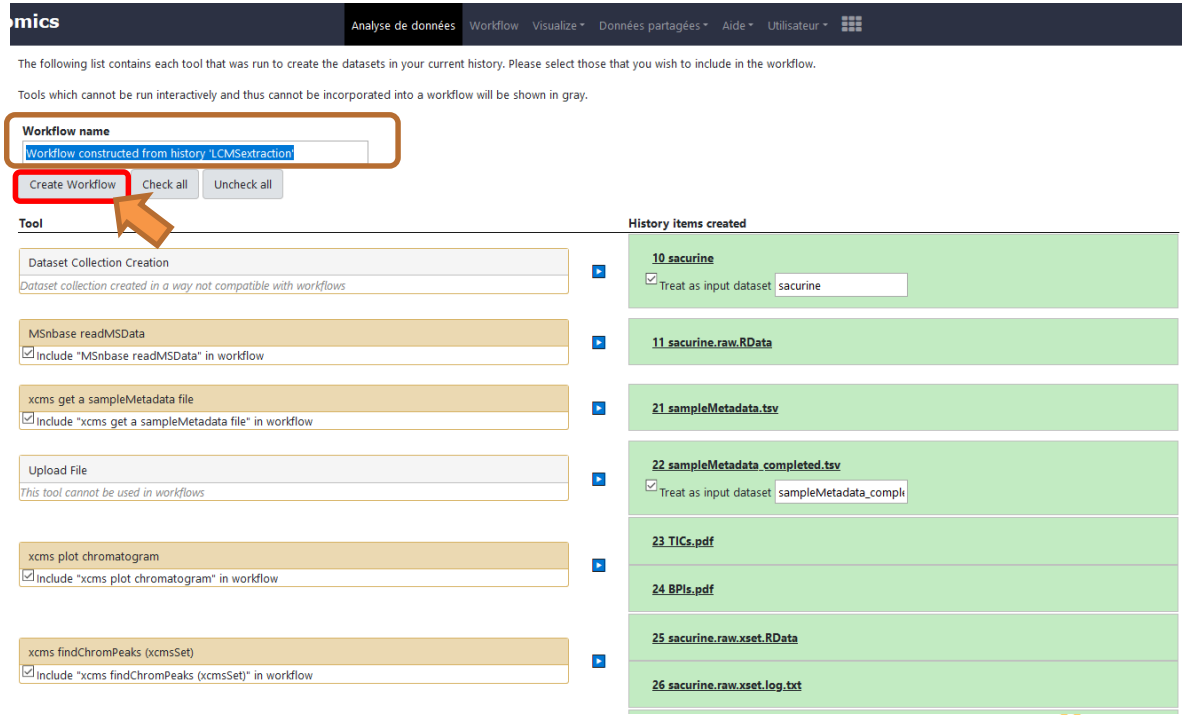
1- History options



2- 'Extract workflow'

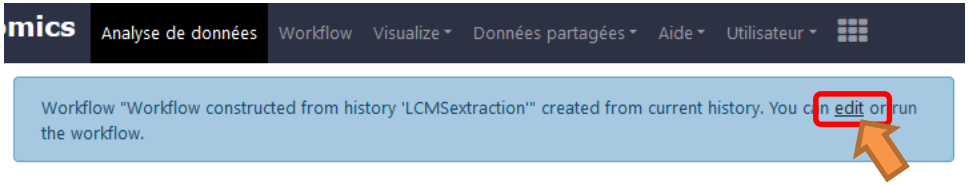
3- Customise your workflow name

4- 'Create Workflow'



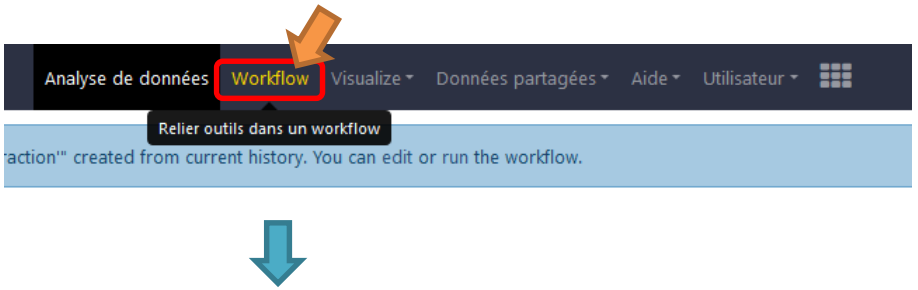
Step 3: Check your workflow (1)

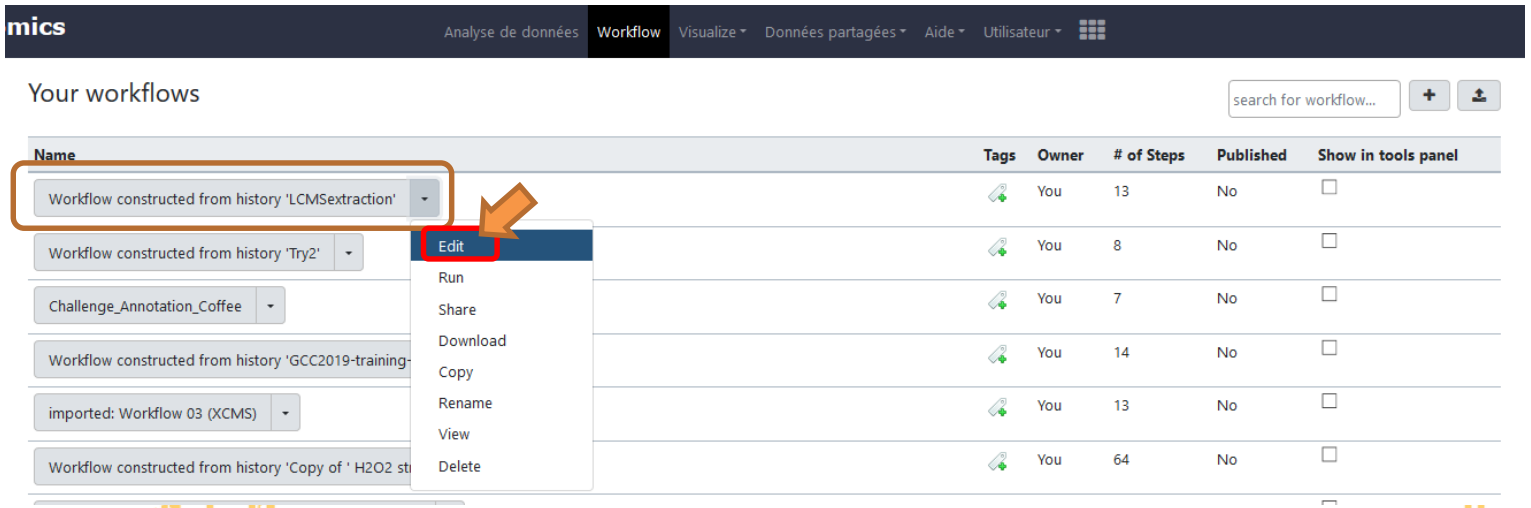
Access the workflow created by clicking on ‘Edit’



OR

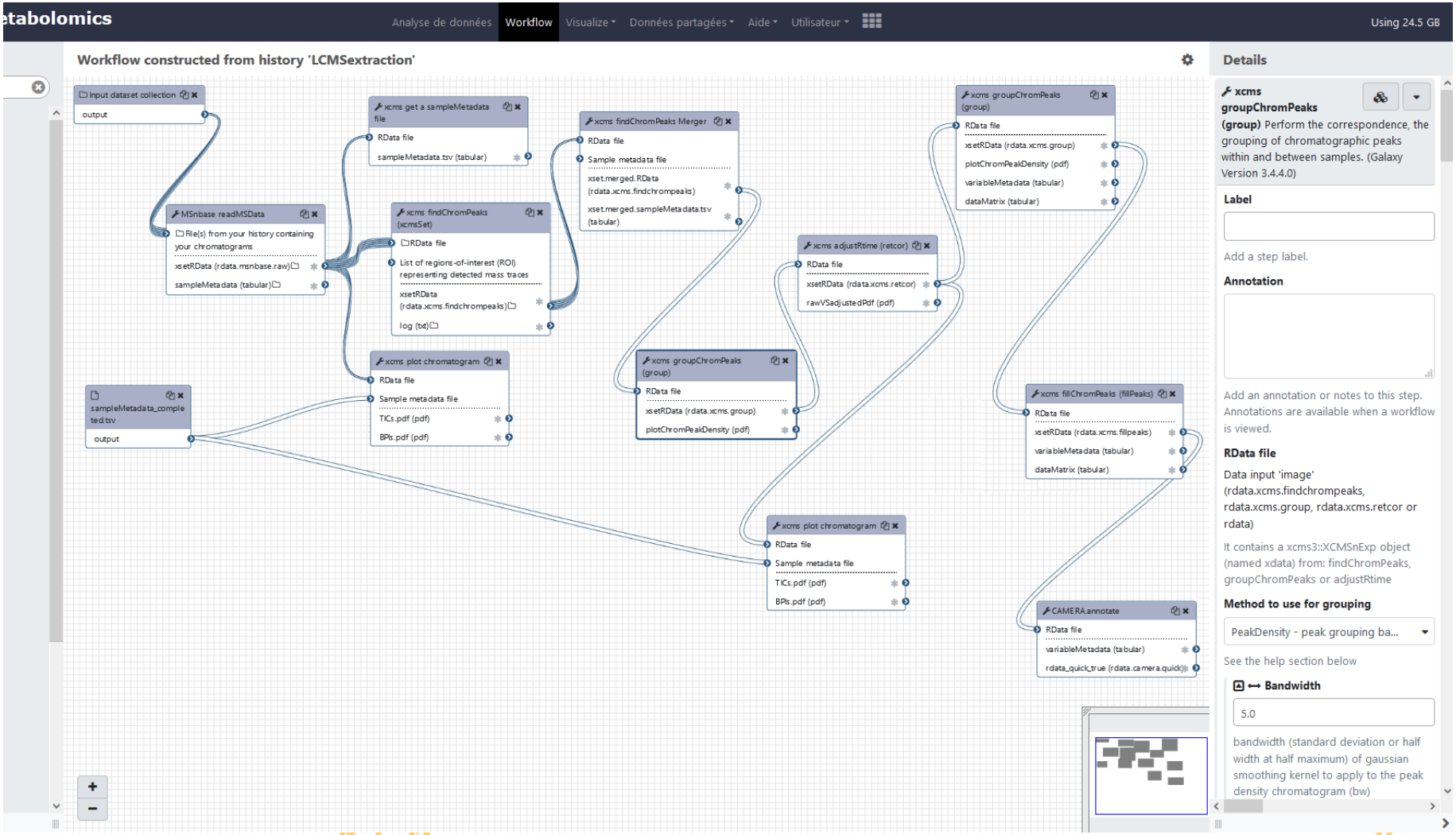
Access the workflow created by accessing the ‘Workflow’ section



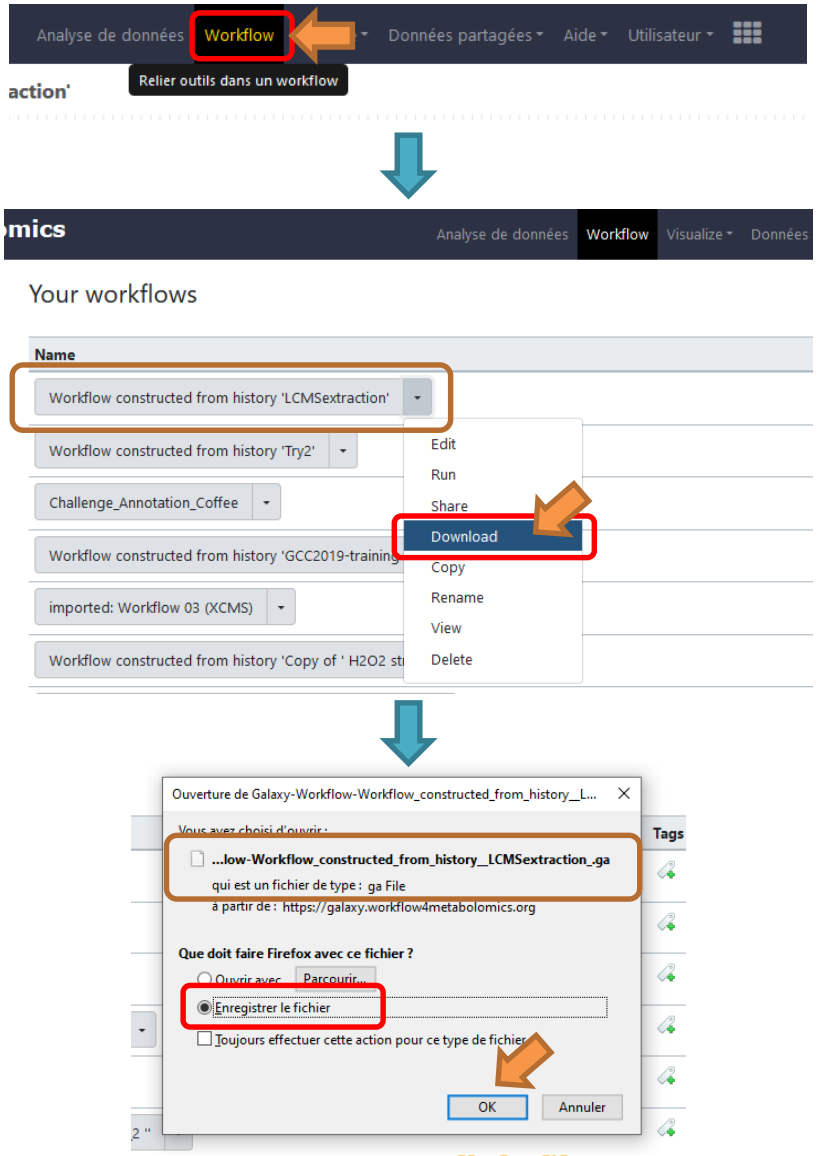


Step 3: Check your workflow (2)

Check that your workflow is complete and no box is alone (at least one link to another box).
In case of anomaly, check your **history** to identify the issue, solve it and extract again.



Step 4: Download your workflow



1- Go back to the 'Workflow' section

2- 'Download' your workflow

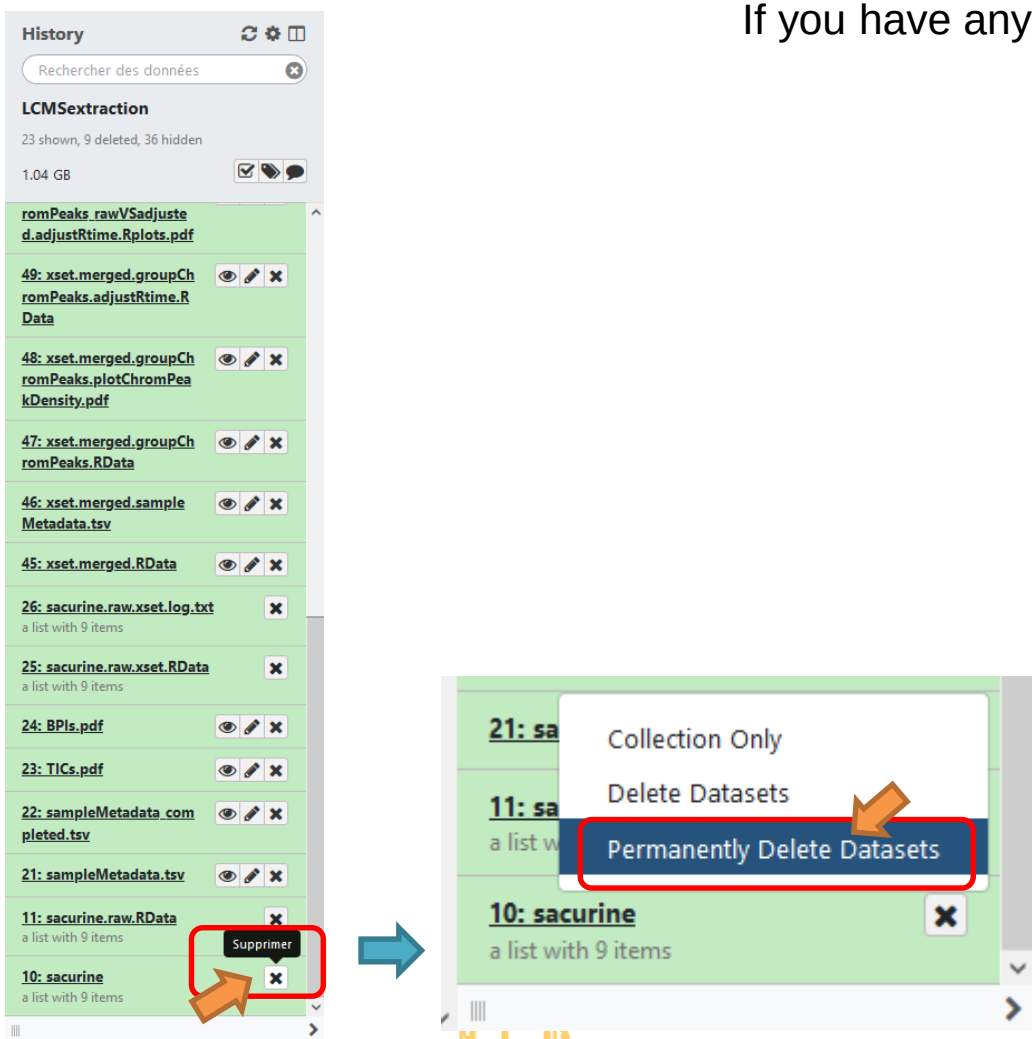
3- Save the file on your computer

This archive file contains the information about the jobs constituting your workflow. Keep it if you need to be able to access the parameters' information in the future.

Step 5: Clean the raw data (if any)

Raw data (*e.g.* '.mzML' files in MS) is data you originally uploaded. It has a large size and you already have these files outside from Galaxy. Keeping it in the history you want to download is irrelevant.

If you have any of this kind of data in your history, delete it.

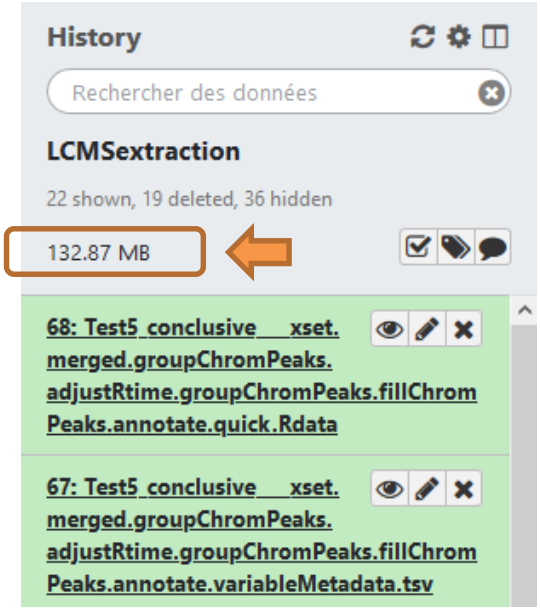


This step deletes **permanently** your dataset collection.

If you are not sure whether your raw data is easily retrievable in your laboratory, check before proceeding.

If necessary, remember that you can download the raw files seperately before doing this cleaning step.

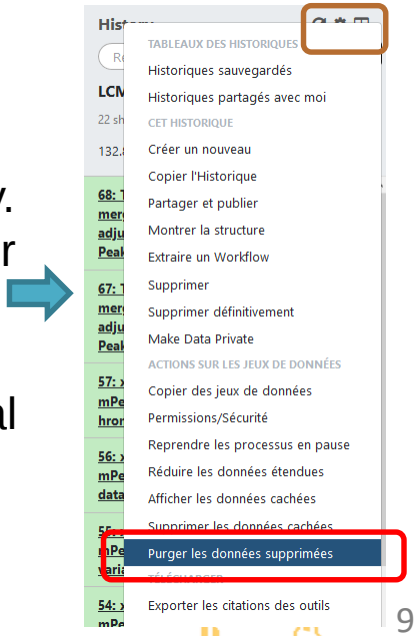
Step 6: Check the size of your history



You are going to download a potentially large file to get you data files out from Galaxy at once. If your history is huge, the download process may be too long.

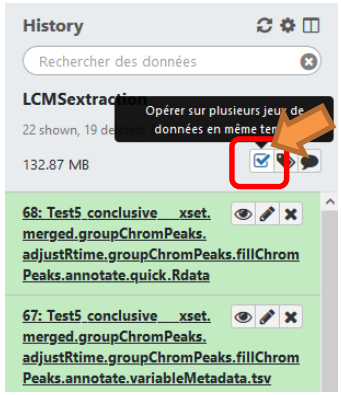
Check the size of your history.

- If the size is < 2 Go: you can proceed to the next step
- If the size is > 2 Go:

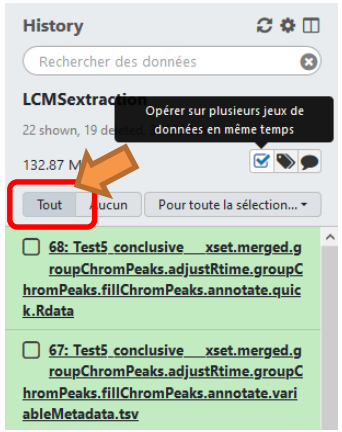


- You may have large files that have been deleted but not permanently. You can choose to purge your deleted datasets and check again your history size. (/!\ deleted datasets will be deleted **permanently**)
- If your history still is too large, you need to 'cut' your history in several parts: define groups of datasets (you can use dataset numbers to help) and proceed to the next step for each of these groups.

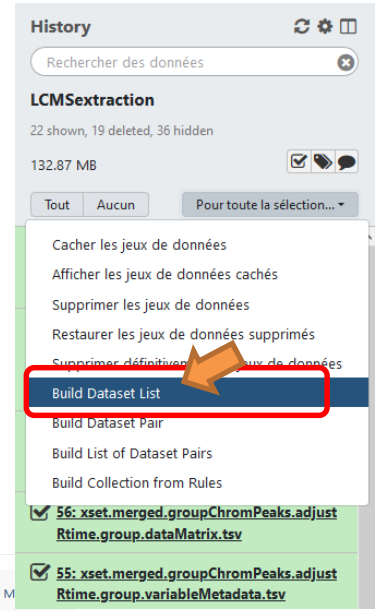
Step 7: Create a dataset collection

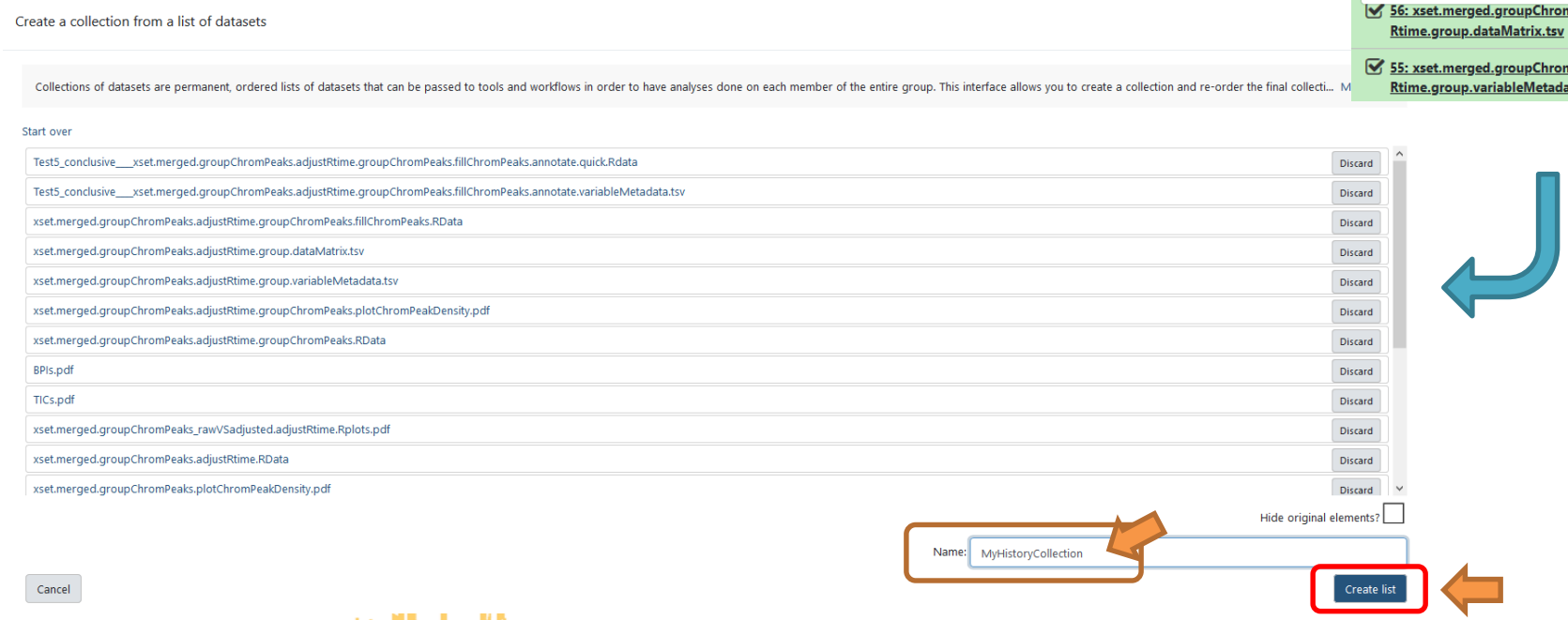


Activate multiple selection



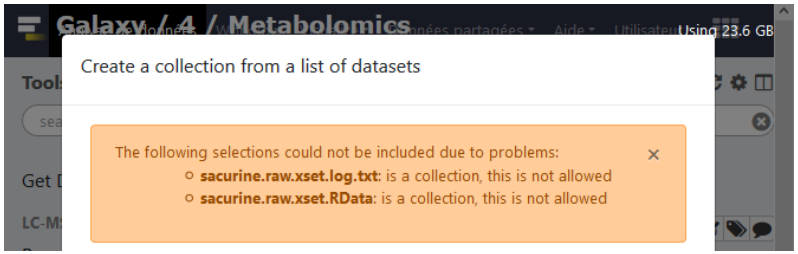
Select all datasets





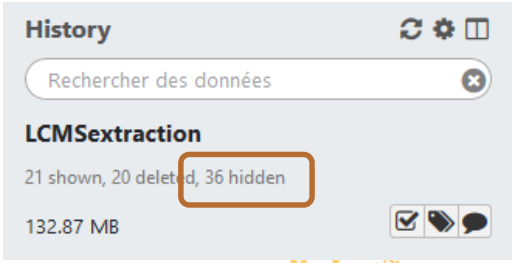
Note: case of already existing dataset collections

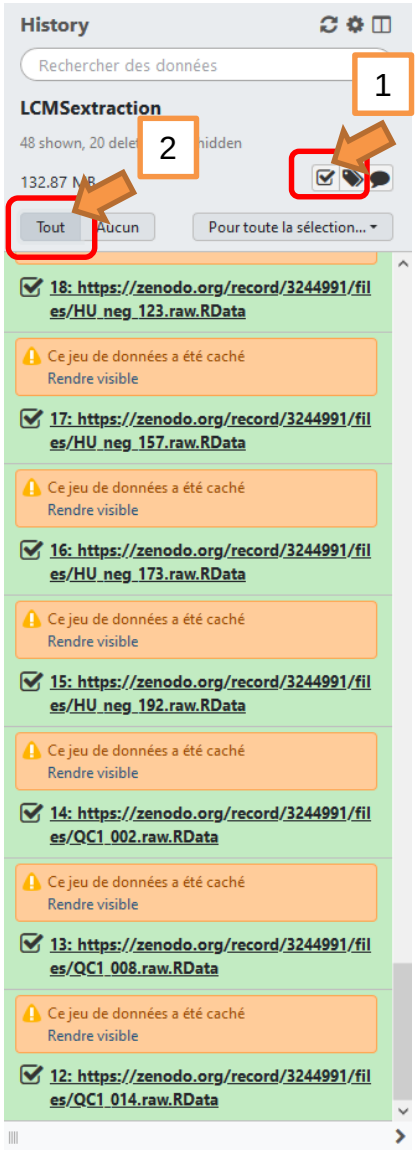
Already existing dataset collections can not be included in another dataset collection.

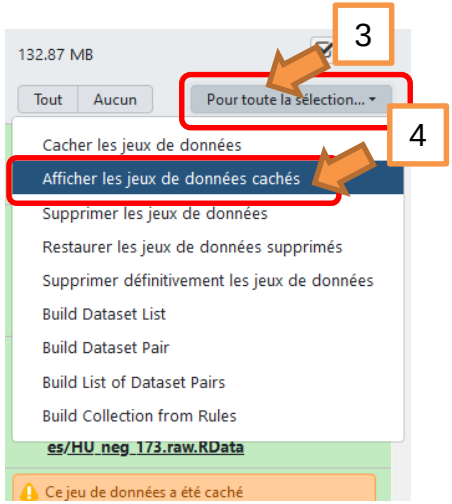


If you already have other dataset collections, you can either:

- download them independently (see next step)
- include corresponding datasets in the previous one by ‘unhiding’ corresponding datasets if hidden



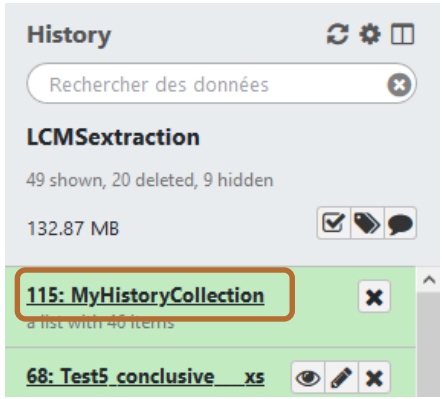


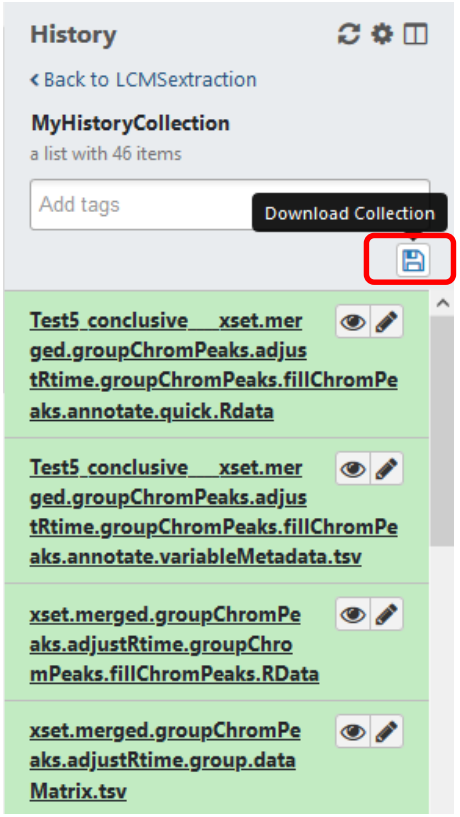


/!\ If you choose the ‘unhiding’ strategy, proceed to it **before** doing step 7

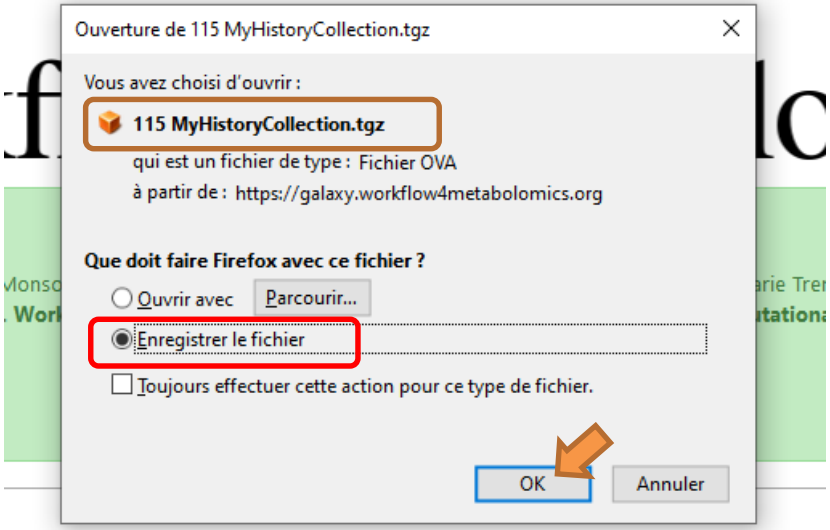
Step 8: Download the dataset collection

Click on the name of your dataset collection





This archive file contains all the files of the dataset collection. You can unzip the archive to get the individual files.



Your downloaded data

Archive containing your history files

Unzip folder of you history files

Archive containing the workflow (information about jobs)

Notes about unzipping your history folder

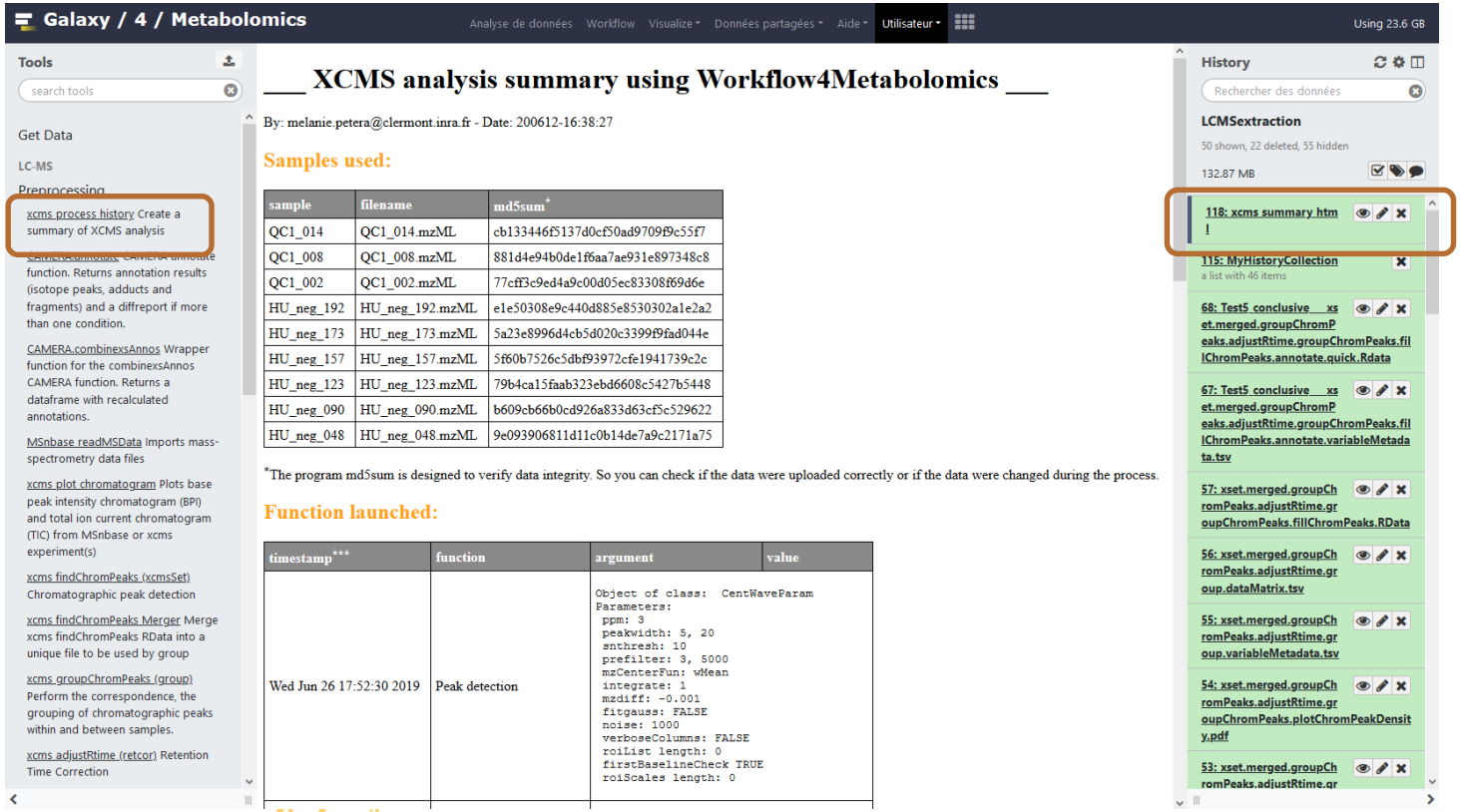
- To unzip the archive obtained by downloading your history dataset collection, you need a dedicated software on your computer.
  If you do not have one, you can download the free software « **7-zip** »
- Please note that the Windows-built-in unzipping solution may not handle correctly this kind of archive.

Notes concerning XCMS histories (1)

Want to gather all your XCMS parameters in a row?

→ Do not forget to run the '*xcms_process_history*' tool on the last step of your XCMS workflow! (*before* constructing your dataset collection to include it in the archive)

Note: If you are interested in keeping a trace of your XCMS parameters only, the HTML file generated by this tool can replace the workflow extraction regarding traceability.



Notes concerning XCMS histories (2)

Most XCMS histories contain dataset collections for the first steps of the extraction process (‘Msnbase_readMSData’, ‘xcms_findChromPeaks_(xcmsSet)’).

As explained previously you can either *download these collections individually* or include them in the overall dataset collection by *unhiding the concerned data files beforehand*.

Please note that these dataset collections contain one file per sample in each collection. This can easily leads to a huge number of files in the archive if you choose the ‘unhiding datasets’ option.

- If your main objective is to download the file in a storage goal, you may consider using the ‘unhiding’ option.
- Otherwise, individual downloads of these collections may be the best option.

Notes concerning XCMS histories (3)

As previously mentioned, raw data (.mzML, .mzXML, .cdf…) is not meant to be downloaded from Galaxy as it was not generated by Galaxy. Since you uploaded it in the first place, you should be able to find the data elsewhere.

If for any reason you still need to download the data from Galaxy, please note that the files' extension you will get will be 'doubled' compared to the original files. For example, a file originally named 'Myrawdata1.mzML' will get que name 'Myrawdata1.mzML.mzml'. Thus, if you want your files to match your sampleMetadata table, you will need to get rid of the added '.mzml' extension first.

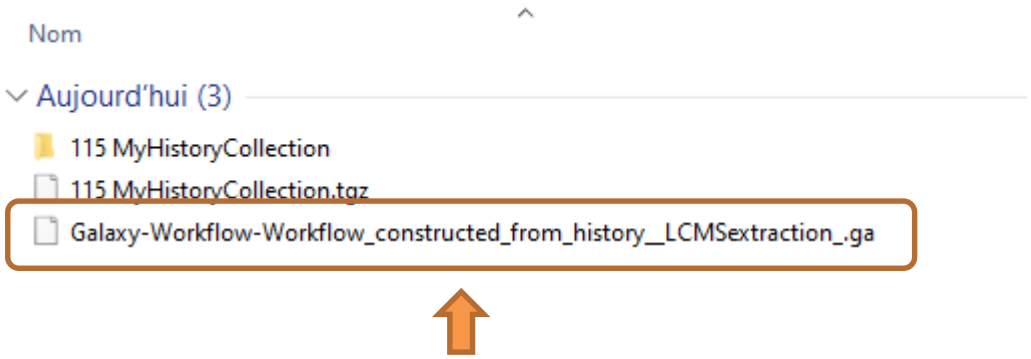
Can't find out where your raw data is? Adopt the Metabolights attitude!

- Such data repositories combine the advantage of (i) storing your data with easy access and DOI and (ii) promoting open-science by going one step further to FAIR.
- Metabolights is the reference data repository in the metabolomics community. Find out more about it here: https://www.ebi.ac.uk/metabolights/
- W4M highly promotes FAIR practices. Tip: in Galaxy we provide a downloader enabling data retrieval directly from Metabolights!

This step is highly recommanded, at least to check that your workflow can be imported smoothly in the new *workflow4metabolomics.usegalaxy.fr* instance

PUTTING YOUR DATA BACK INTO GALAXY

Step 1: the workflow archive



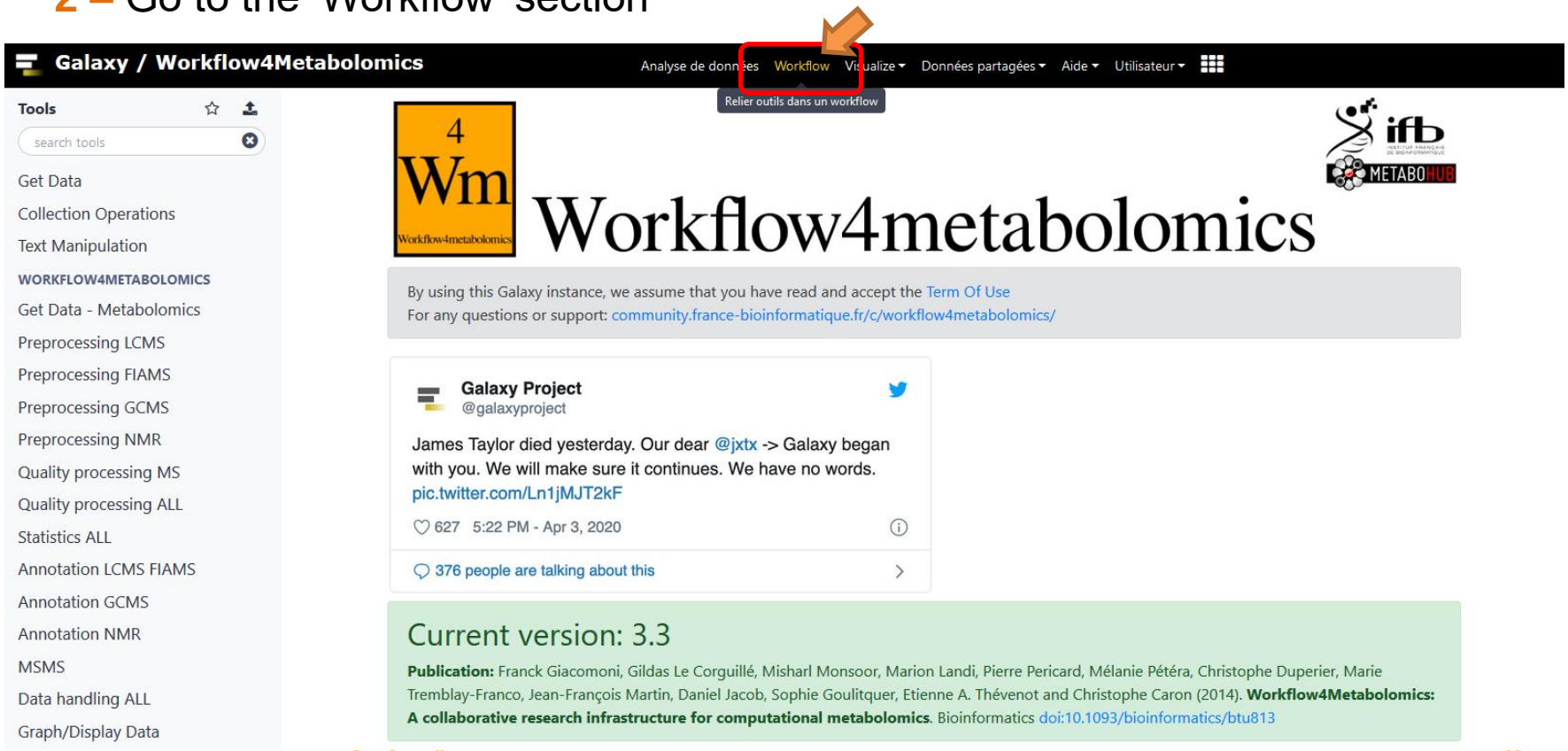
Make sure you have your workflow archive on your computer

Step 2: Go to the 'Workflow' section

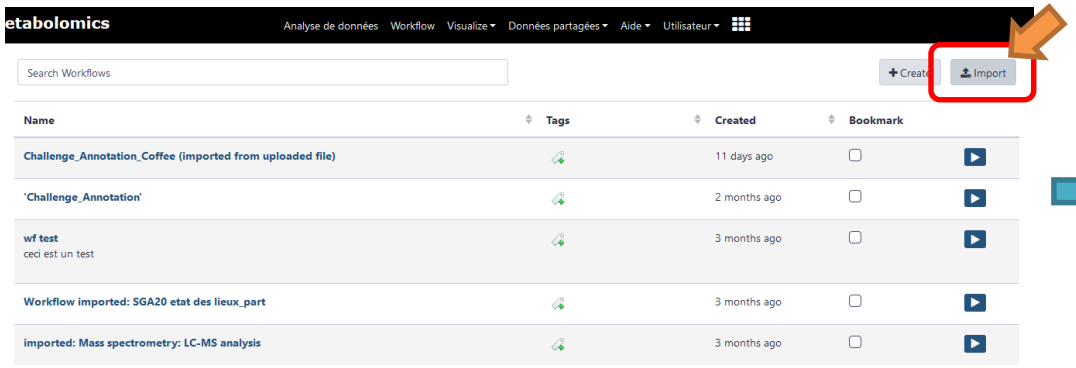
1 - Log in workflow4metabolomics.usegalaxy.fr

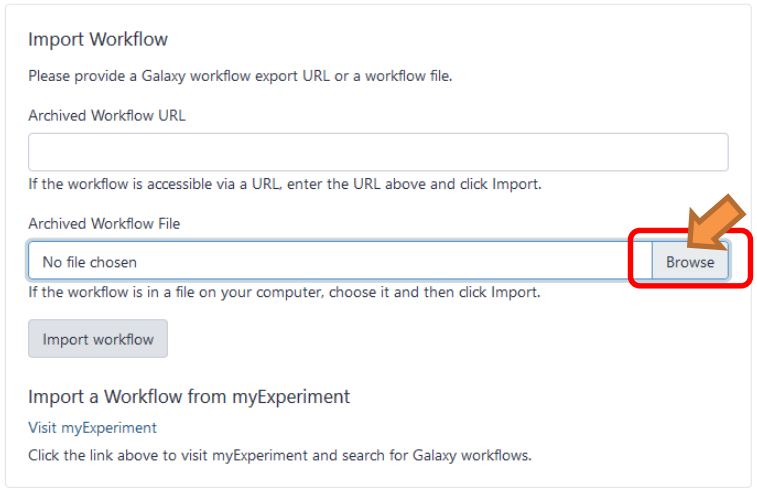
If you have not got an account yet, you can create one (auto-registration)

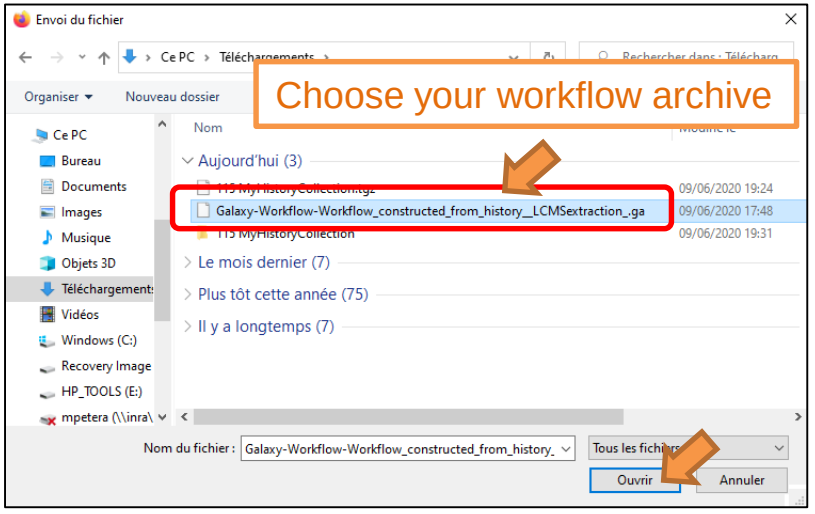
2 – Go to the 'Workflow' section

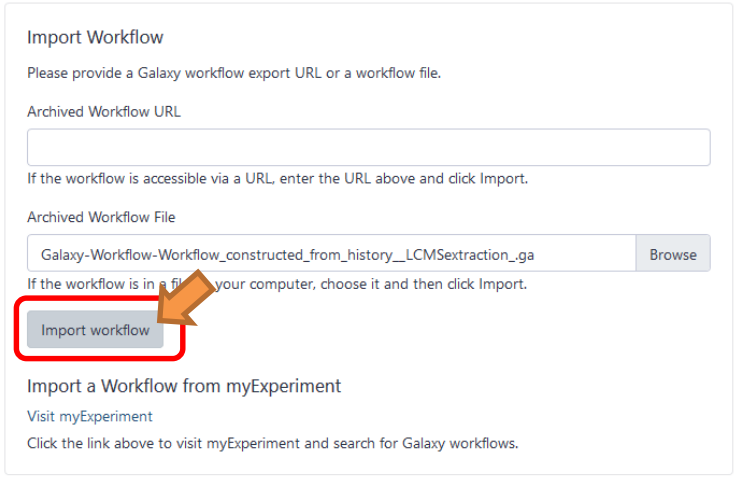


Step 3: Import your workflow

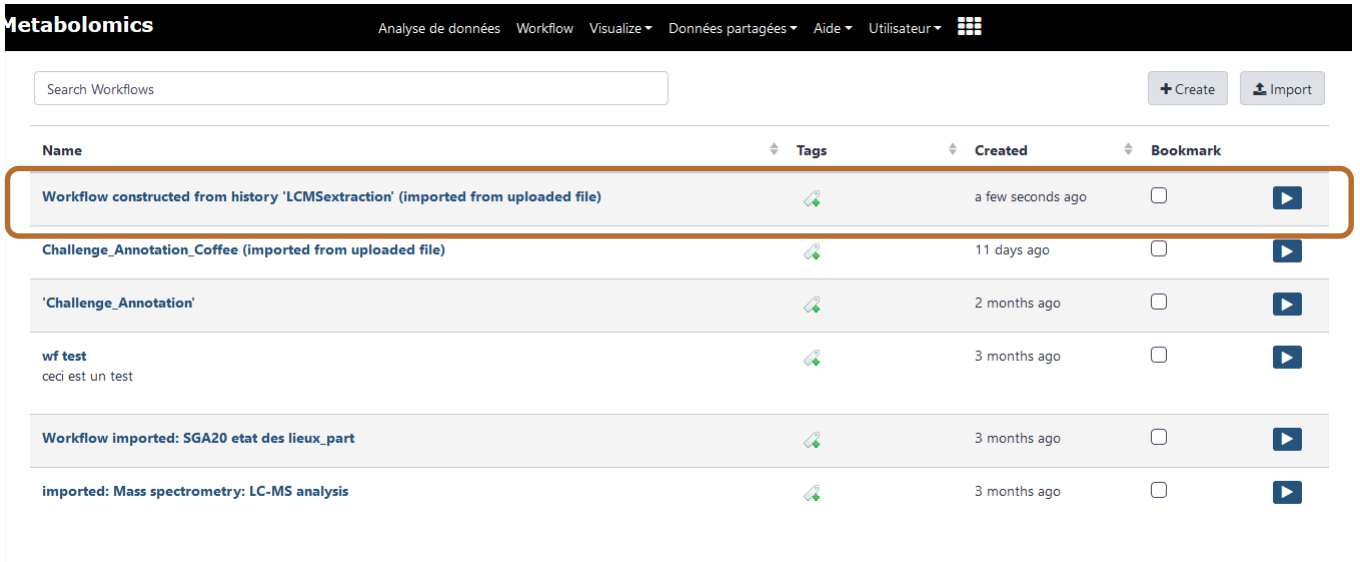








Your workflow is ready to be used!



You can either:

- View it to get access to the jobs and parameters you used
- Use it to relaunch your analysis – in a new history, simply upload from your computer the input data needed, then run the workflow on these data

Workflow troubleshooting (1)

'I have never used workflows before. How do I recreate my history in the new instance?'

As mentioned previously, all you need to do is to upload the input data needed for your workflow (*e.g.* in case of XCMS your raw data and the completed sampleMetadata file) in a new history, and then to launch the workflow using it.

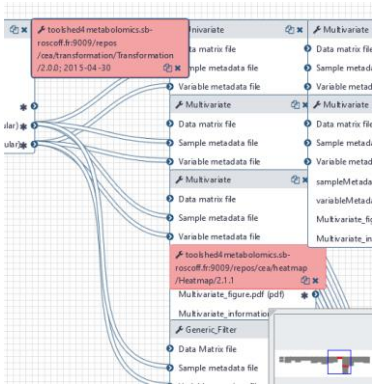
If you are not familiar with how to launch a workflow, you can find out how following the 'Galaxy 101' GTN tutorial:
https://galaxyproject.github.io/training-material/topics/introduction/tutorials/galaxy-intro-101-everyone/tutorial.html
You can focus on the '*Galaxy management*' chapter.

Workflow troubleshooting (2)

'It seems I can not import all my workflow in the new galaxy instance… What should I do?'

There are two configurations that you might run into while facing workflow import issues:

- The tool is missing – you get 'red boxes' in your workflow



While trying running it:

> Following tools missing: toolshed4metabolomics.sb-roscoff.fr:9009/repos/cea/transformation/Transformation/2.0.0; 2015-04-30, toolshed4metabolomics.sb-roscoff.fr:9009/repos/cea/heatmap/Heatmap/2.1.1

While trying editing it:

- Step 13: toolshed4metabolomics.sb-roscoff.fr:9009/repos/cea/transformation/Transformation /2.0.0; 2015-04-30
  - Tool is not installed

- The tool exists however the version that have been used is not available – the tool box is displayed in your workflow but with default parameter values

While trying running it:

> Some tools in this workflow may have changed since it was last saved or some errors were found. The workflow may still run, but any new options will have default values. Please review the messages below to make a decision about whether the changes will affect your analysis.

While trying editing it:

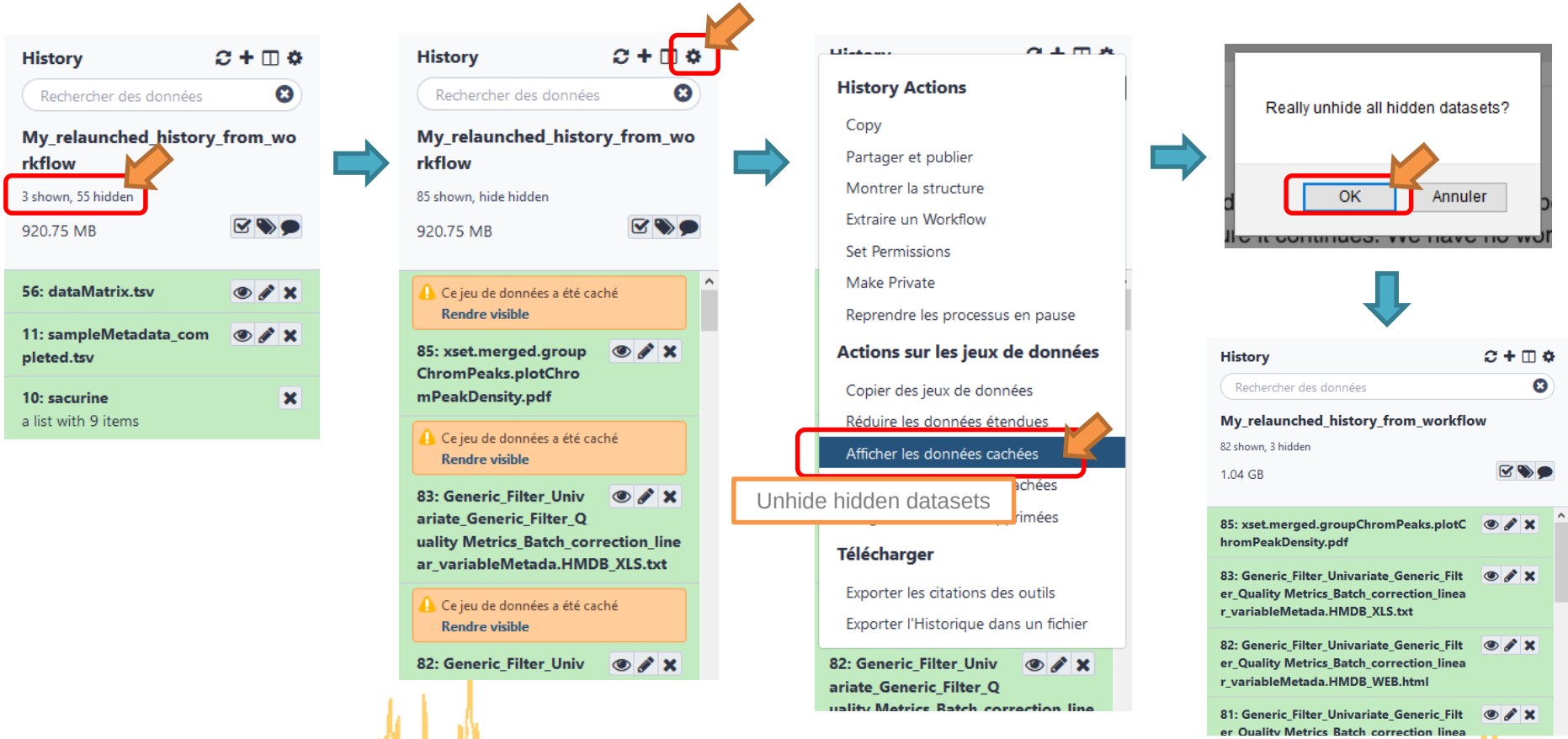
- Step 7: xcms fillChromPeaks (fillPeaks)
  - No value found for 'Convert retention time (seconds) into minutes'. Using default: 'False'.
  - No value found for 'Reported intensity values'. Using default: 'into'.

If you run into one of these issues, please visit the "Troubleshooting – Workflows" post on the IFB forum for more information
https://community.france-bioinformatique.fr/t/troubleshooting-workflows/615

Workflow troubleshooting (3)

'I relaunched my workflow but I do not see resulting jobs in my history. What should I do?'

It sometimes happens that while using imported workflows, job outputs appear as 'hidden datasets' in your history. To put it back to standard display, you simply need to select concerned datasets and to unhide them.



See you on W4M!

Website:
https://workflow4metabolomics.org/

Galaxy instance:
https://workflow4metabolomics.usegalaxy.fr/

Need some help?
https://community.france-bioinformatique.fr/c/workflow4metabolomics/10